



Review

# Accuracy of Machine Learning Assisted Detection of Keratoconus: A Systematic Review and Meta-Analysis

Ke Cao <sup>1,2</sup>, Karin Verspoor <sup>3,4</sup>, Srujana Sahebjada <sup>1,2,†</sup> and Paul N. Baird <sup>2,\*,†</sup>

- <sup>1</sup> Centre for Eye Research Australia, Melbourne, VIC 3002, Australia; kcao1@student.unimelb.edu.au (K.C.); srujana.sahebjada@unimelb.edu.au (S.S.)  
<sup>2</sup> Department of Surgery, Ophthalmology, The University of Melbourne, Melbourne, VIC 3002, Australia  
<sup>3</sup> School of Computing Technologies, RMIT University, Melbourne, VIC 3000, Australia; karin.verspoor@rmit.edu.au  
<sup>4</sup> School of Computing and Information Systems, The University of Melbourne, Melbourne, VIC 3010, Australia  
\* Correspondence: pbaird@unimelb.edu.au; Tel.: +61-3-9929-8613  
† These authors contributed equally as senior authors.

**Abstract:** (1) Background: The objective of this review was to synthesize available data on the use of machine learning to evaluate its accuracy (as determined by pooled sensitivity and specificity) in detecting keratoconus (KC), and measure reporting completeness of machine learning models in KC based on TRIPOD (the transparent reporting of multivariable prediction models for individual prognosis or diagnosis) statement. (2) Methods: Two independent reviewers searched the electronic databases for all potential articles on machine learning and KC published prior to 2021. The TRIPOD 29-item checklist was used to evaluate the adherence to reporting guidelines of the studies, and the adherence rate to each item was computed. We conducted a meta-analysis to determine the pooled sensitivity and specificity of machine learning models for detecting KC. (3) Results: Thirty-five studies were included in this review. Thirty studies evaluated machine learning models for detecting KC eyes from controls and 14 studies evaluated machine learning models for detecting early KC eyes from controls. The pooled sensitivity for detecting KC was 0.970 (95% CI 0.949–0.982), with a pooled specificity of 0.985 (95% CI 0.971–0.993), whereas the pooled sensitivity of detecting early KC was 0.882 (95% CI 0.822–0.923), with a pooled specificity of 0.947 (95% CI 0.914–0.967). Between 3% and 48% of TRIPOD items were adhered to in studies, and the average (median) adherence rate for a single TRIPOD item was 23% across all studies. (4) Conclusions: Application of machine learning model has the potential to make the diagnosis and monitoring of KC more efficient, resulting in reduced vision loss to the patients. This review provides current information on the machine learning models that have been developed for detecting KC and early KC. Presently, the machine learning models performed poorly in identifying early KC from control eyes and many of these research studies did not follow established reporting standards, thus resulting in the failure of these clinical translation of these machine learning models. We present possible approaches for future studies for improvement in studies related to both KC and early KC models to more efficiently and widely utilize machine learning models for diagnostic process.

**Keywords:** keratoconus; diagnosis; early detection; artificial intelligence; machine learning; reporting completeness



**Citation:** Cao, K.; Verspoor, K.; Sahebjada, S.; Baird, P.N. Accuracy of Machine Learning Assisted Detection of Keratoconus: A Systematic Review and Meta-Analysis. *J. Clin. Med.* **2022**, *11*, 478. <https://doi.org/10.3390/jcm11030478>

Academic Editor: Hidenaga Kobashi

Received: 29 November 2021

Accepted: 13 January 2022

Published: 18 January 2022

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Corneal diseases are the second largest cause of blindness worldwide, behind only cataract in overall importance [1]. Keratoconus (KC), one of the most common corneal conditions, is characterized by bilateral, progressive corneal thinning that results in an abnormally steep cornea, and decreased vision [2]. The disease primarily affects young adults and children [3]. Globally KC patients comprise the second largest group of patients

requiring corneal transplants, and there are thus associated risks of surgical complications and immune rejection of the graft [4,5].

In the medical field, the recent availability of biomedical data has led to the advent of the big data era [6,7], creating opportunities for more comprehensive data-informed decision making. The challenge for the clinician has expanded beyond data collection to encompass the interpretation of a greater amount of information. Artificial intelligence (AI) is the ability of a machine to learn and display intelligence [8]. Machine learning methods represent a branch of AI where computational algorithms can be used to process and identify patterns in large amounts of data at a scale that is beyond the ability of humans to synthesize. Through advanced pattern mining, innovative detection solutions, referred to as automatic detection models, can be based on these massive amounts of data [9].

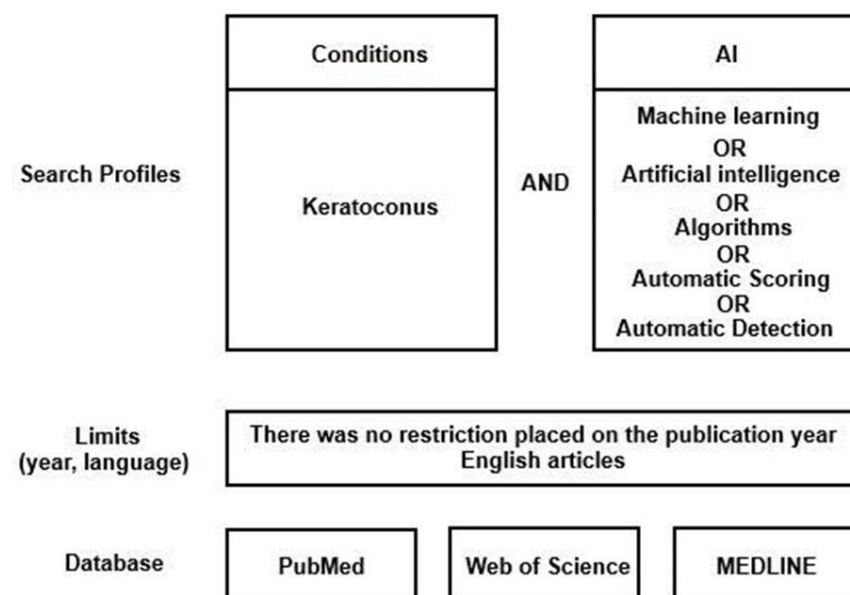
Increased detailed data about the cornea resulting from corneal topography and tomography systems are essential for diagnosing KC. They are especially useful for detecting early indications of KC, prior to the development of typical KC clinical characteristics [10]. Corneal tomography provides parameters and images [11], which are laborious to analyze manually. This has also been a motivating force in the use of machine learning for KC detection since 1995 [12], with a growing volume of machine learning research being conducted in KC detection over the following years.

There is no existing study that summarizes the use of machine learning in KC, identifies limitations, and makes recommendations for future directions. The aim of this study and meta-analysis was to systematically review all currently available literature to determine accuracy (through the use of pooled sensitivity and specificity) of machine learning in the detection of KC, addressing this knowledge gap.

## 2. Materials and Methods

### 2.1. Literature Search Strategy

A web-based systematic literature search was performed for articles published from inception through 28 February 2021, on PubMed, Web of Science, and MEDLINE (Figure 1). Database searches were supplemented by hand-search and grey literature search techniques to ensure all publications in this field were included. The protocol for this systematic review was registered on PROSPERO (registration number CRD42021237167).



**Figure 1.** The search strategy used in the present study.

PubMed, Web of Science, and MEDLINE were interrogated using search strings pertaining to keratoconus and machine learning. Key words were used by search engine

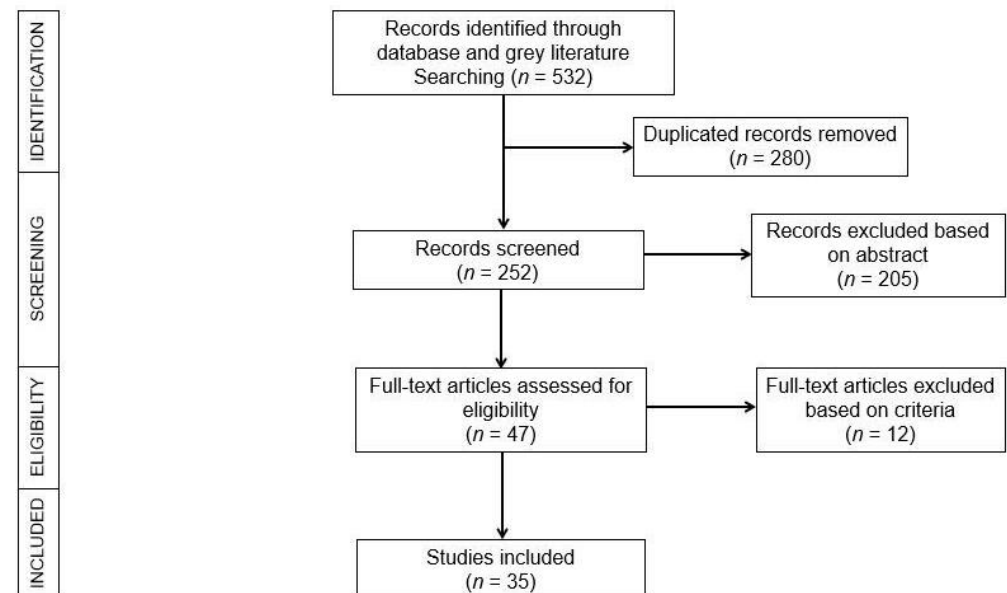
and designated filters according to PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines (Figure 1). The search string was ((keratoconus AND [artificial intelligence]) OR (keratoconus AND [machine learning]) OR (keratoconus AND algorithms) OR (keratoconus AND [automatic scoring]) OR (keratoconus AND [automatic detection])).

The following sections define the inclusion and exclusion criteria.

Primary research articles, meeting the following requirements were deemed suitable for inclusion in this review:

- Full text original papers that evaluated machine learning algorithms in the diagnosis of KC;
- No limit on the year of publication was applied;
- Publications only in the English language were included;
- Publications in which KC was the only corneal condition evaluated.
- Criteria for exclusion:
  - Publications evaluating other corneal diseases without focusing only on KC;
  - Publications evaluating the efficacy of machine learning in treating KC (treatments);
  - Review papers;
  - Publications in which no machine learning algorithms were included, but only statistical research was undertaken;
  - Non-English publications.

The review articles were imported into Endnote (version X9), which was used to perform the initial screening. Two independent reviewers (K.C. and S.S.) screened the titles and abstracts of journals for possible qualifying studies, and inconsistencies were settled by consensus. Both reviewers thoroughly analyzed all potential qualifying studies for inclusion/exclusion. (Figure 2).



**Figure 2.** The PRISMA flowchart illustrating the literature selection process.

## 2.2. Data Extraction

A customized analysis form was generated in Microsoft Excel. The following items were reviewed from each study: authors, publication year, country, study objective, sample size, reference standard diagnosis methods, corneal imaging systems used to generate data and machine learning method used, indicators of studies' quality, and the number of true positives (TP), false positives (FP), true negatives (TN), and false negatives (FN). If the number of TP, FP, TN, or FN were not presented, then these values were derived from the data provided, such as sensitivity and specificity. The meta-analysis omitted studies that

lacked data on TP, FP, TN, or FN or that lacked the capacity to measure these data to create a two-way contingency chart.

### 2.3. Reporting Completeness of Machine Learning Studies in KC

We evaluated the reporting completeness of machine learning research in this study by referring to the TRIPOD (transparent reporting of a multivariable prediction model for individual prognosis or diagnosis; [www.tripod-statement.org](http://www.tripod-statement.org) (accessed on 7 June 2021)) statement relevant to model development. This statement contains a 20-item checklist, totaling 31 items when all sub-items are included. The checklist includes questions about the title, abstract, background, methods, results, discussion, supplementary material, and funding information. Two items (5c, i.e., “Give details of treatments received, if relevant”, and 11, i.e., “Provide details on how risk groups were created, if done”) were omitted since they were irrelevant to the research covered in this review. Each study was therefore evaluated on a total of either 28 or 29 possible items. This total number of items varied between 28 and 29 since item 14b (i.e., “If nothing on univariable analysis (in methods or results) is reported, score not applicable”) may be rated as “not applicable” and was thus omitted from the calculation of reporting adherence. If a study had data for several models, we extracted data for the model with the highest performance. Each included item received a score of “1” for adherence and a score of “0” for non-adherence. Multiple items (items 1, 2, 3a, 4b, 5a, 6a, 7a, 7b, 9, 10a, 10b, 10d, 13a, 13b, 14a, 15a, 16, 17, 20 and 22) in the TRIPOD analysis were derived from several sub-items (the sub-items for each number can be found in [www.tripod-statement.org](http://www.tripod-statement.org) (accessed on 7 June 2021)). The score was therefore determined by the combination of several elements rather than a single element. The results of each TRIPOD item for each paper and the level of reporting adherence for each TRIPOD item were documented systematically in a spreadsheet.

For each machine learning study, we calculated the TRIPOD adherence score by dividing the sum of TRIPOD items adhered to by the entire number of applicable TRIPOD items in the study. The average adherence score was calculated using the median value of the adherence score across all studies. For each TRIPOD item, the adherence score was calculated by dividing the number of studies that adhered to the item by the total number of applicable studies for the item. The median value was used to represent the average adherence score for each TRIPOD item.

### 2.4. Statistical Methods

All analyses were configured in ‘mada’ and ‘metafor’ package from RStudio Server Pro (PBC, Boston, MA) (Version 1.3.1056-1) for Windows. To measure the overall machine learning performance for KC detection, the sensitivity and specificity values for all presented models were pooled, following the bivariate meta-analysis method of Reitsma et al. [13] using linear mixed model techniques. The bivariate method retains the two-dimensional nature (sensitivity and specificity) of the original data. Pairs of sensitivity and specificity are jointly evaluated, accommodating any correlation that could exist between sensitivity and specificity using a random effects approach. This was accomplished using the ‘mada’ package (Version 0.5.10) from RStudio. The 95% confidence interval (CI) of the sensitivity and specificity of various imaging systems were compared. The Deeks’ funnel asymmetry test, developed especially for diagnostic test accuracy [14], was used to determine if there was cross-study publishing bias [15].

A hierarchical summary receiver-operating characteristic (HSROC) curve was fitted. Each individual study was presented as a circle and plotted within the HSROC curve. The summary point was represented by a dot surrounded by a 95% confidence interval (95% CI).

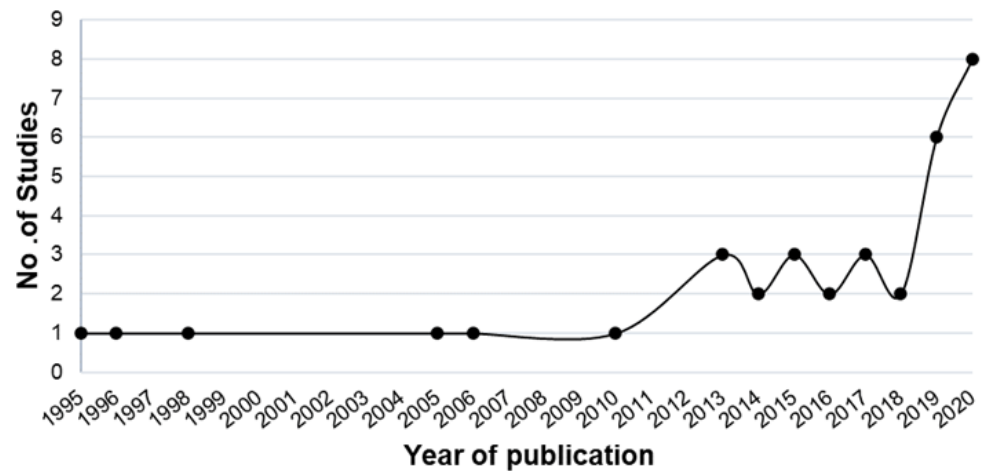
### 2.5. Outcomes Measure

The primary outcome indicator was the diagnostic accuracy of machine learning algorithms through a variety of imaging technologies for the identification of KC, as determined by the pooled sensitivity and specificity values.

## 3. Results

### 3.1. Search Collection

Of the initial literature search, 532 studies were retrieved, and 280 duplicates were omitted. Following review of the title, abstract, and full text, 35 studies were included in the review. There was an increasing trend of machine learning studies in KC published over time (Figure 3), with the earliest study published in 1995 by Maeda et al. [12] and increasing to over 30 studies in 2020.



**Figure 3.** The count of machine learning studies in KC from 1995 to 2020.

### 3.2. Search Characteristics

The 35 articles on machine learning and KC were reviewed were classified into three categories based on their aims: detecting KC eyes from controls, differentiating early KC from controls, and identifying different KC severities. Each study focused on one or more of these aims, with 12 research papers focused on KC versus control, 4 research papers focused on early KC versus control, 1 article focused on KC severity, and 18 publications with multiple aims. There was no study examining the progression of KC.

Following classification of studies with multiple aims, the following results were obtained: 30 studies [12,16–44] have evaluated machine learning models to distinguish KC from controls, 14 studies [20,21,27–29,31,33,37,39,44–48] have evaluated machine learning models to distinguish early KC from controls, and 6 studies [12,33,35,40,42,49] have assessed machine learning models in KC staging. In this review, we utilize the term early KC rather than subclinical KC or forme fruste KC, due to a lack of unified criteria for these terms [20,21,27–29,31,33,37,39,44–50]. Meta-analyses were conducted on each group of studies. A summary of the final 35 studies included in this current study can be found in Table 1.

**Table 1.** Identified studies using machine learning in detection of KC and early KC.

Study Objectives	First Author	Year	No. of Centers Involved (Country)	Sample Size	No. of KC/Early KC Eyes	Machine Learning Method/s Used	Data Type (No. of Parameters)	Corneal Imaging Modality	Evaluation Methods
Detect KC eyes from controls	Maeda et al. [12]	1995	1 (USA)	176	44	Combined discriminant analysis and classification tree	P (8)	TMS-1	Internal
	Kalin et al. [16]	1996	NR	106	5	Combined discriminant analysis and classification tree	P (8)	TMS-1	Validation study
	Rabinowitz et al. [17]	1998	1 (USA)	241	99	Linear discriminant analysis	P (5)	TMS-1	Internal
	Twa et al. [18]	2005	NR (USA)	244	112	Decision tree	P (36)	Keratron	Internal
	Bessho et al. [19]	2006	2 (Japan)	165	63	logistic regression	P (na)	Orbscan II	External
	Saad et al. [20]	2010	NR	143	31	Discriminant analysis	P (51)	Orbscan IIz	Internal
	Smadja et al. [21]	2013	1 (France)	325	148	Decision tree	P (55)	GALILEI	Internal
	Mahmoud et al. [22]	2013	3 (Colombia, USA, Switzerland)	407	163	logistic regression	P (na)	GALILEI	External
	Saad et al. [23]	2014	1 (France)	166	64	Discriminant analysis	P (7)	Orbscan IIz	Internal
	Silverman et al. [24]	2014	1 (UK)	204	74	Multiple methods	P (161)	Artemis-1	Internal
	Koprowski et al. [25]	2015	1 (Brazil)	746	477	Decision tree	P (11)	Corvis	Internal
	Shetty et al. [26]	2015	1 (India)	128	85	Logistic regression	P (na)	Pentacam	Internal
	Kovacs et al. [27]	2016	1 (Hungary)	120	60	Neural network	P (na)	Pentacam HR	Internal
	Ruiz et al. [28]	2016	1 (Belgium)	648	454	Support vector machine	P (22)	Pentacam HR	Internal
	Ambrosio et al. [29]	2017	2 (Brazil, Italy)	756	276	Multiple methods	P (na)	Pentacam HR & Corvis ST	Internal
	Silverman et al. [30]	2017	1 (USA)	141	30	Discriminant analysis	P (240)	Artemis-1 & Pentacam	Internal
	Lopes et al. [31]	2018	5 (UK, Brazil, Italy, USA)	3648	370	Multiple methods	P (na)	Pentacam	Internal & External
	Chandapura et al. [32]	2019	NR	439	218 including 102 early KC	Random forest	P (27)	Pentacam & OCT	Internal
	* Dos Santos et al. [41]	2019	1 (Austria)	142	70	Convolutional neural network	I	OCT	Internal
	Issarti et al. [33]	2019	1 (Belgium)	624	312	Neural network	P (28)	Pentacam	Internal

Table 1. Cont.

Study Objectives	First Author	Year	No. of Centers Involved (Country)	Sample Size	No. of KC/Early KC Eyes	Machine Learning Method/s Used	Data Type (No. of Parameters)	Corneal Imaging Modality	Evaluation Methods
Detect early KC eyes from controls	Kamiya et al. [42]	2019	1 (Japan)	543	304	Convolutional neural network	I	AS-OCT	Internal
	* Lavric et al. [43]	2019	NR	3000	1500	Convolutional neural network	I	SyntEyes model	Internal
	Leão et al. [34]	2019	2 (Brazil, Italy)	574	223	Discriminant analysis	P (na)	Corvis ST	NR
	Bolarin et al. [35]	2020	1 (Spain)	169	107	logistic regression	P	Sirius	Internal
	Castro-Luna et al. [36]	2020	1 (Spain)	60	30	Naive Bayes	P	CSO	Internal
	* Issarti et al. [37]	2020	2 (Belgium)	812	508	Neural Network	P (90)	Pentacam HR	Internal & External
	Kuo et al. [44]	2020	1 (Taiwan)	326	170	Convolutional neural network	I	TMS-4	Internal
	Lavric et al. [38]	2020	NR	3151	1181 including 791 early KC	Multiple methods	P (443)	SS-1000 CASIA OCT	Internal
	* Shi et al. [39]	2020	1 (China)	121	38	Neural network	P (49)	UHR-OCT & Pentacam HR	Internal
	Velazquez-Blazquez et al. [40]	2020	1 (Spain)	178	104 including 61 early KC	Logistic regression	P (27)	Sirius	Internal
	Saad et al. [20]	2010	NR	143	40	Discriminant analysis	P (51)	Orbscan IIZ	Internal
	Smadja et al. [21]	2013	1 (France)	224	47	Decision tree	P (55)	GALILEI	Internal
	* Ventura et al. [45]	2013	NR (Brazil)	204	68	Neural network	P (41)	Ocular Response Analyzer	Internal
	Chan et al. [46]	2015	1 (Singapore)	128	24	Discriminant analysis	P (na)	Orbscan IIZ	Validation study
Kovacs et al. [27]	2016	1 (Hungary)	75	15	Neural network	P (na)	Pentacam HR	Internal	
Ruiz et al. [28]	2016	1 (Belgium)	261	67	Support vector machine	P (22)	Pentacam HR	Internal	
Ambrosio et al. [29]	2017	2 (Brazil, Italy)	574	94	Multiple methods	P (na)	Pentacam HR & Corvis ST	Internal	
Xu et al. [47]	2017	1 (China)	363	77	Discriminant analysis	P (na)	Pentacam HR	Internal	
Lopes et al. [31]	2018	5 (UK, Brazil, Italy, USA)	3537	259	Multiple methods	P (na)	Pentacam	Internal & External	
Issarti et al. [33]	2019	1 (Belgium)	389	77	Neural network	P (28)	Pentacam	Internal	

Table 1. Cont.

Study Objectives	First Author	Year	No. of Centers Involved (Country)	Sample Size	No. of KC/Early KC Eyes	Machine Learning Method/s Used	Data Type (No. of Parameters)	Corneal Imaging Modality	Evaluation Methods
	Cao et al. [48]	2020	1 (Australia)	88	49	Multiple methods	P (11)	Pentacam	Internal
	* Issarti et al. [37]	2020	2 (Belgium)	812	117	Neural Network	P (90)	Pentacam HR	Internal & External
	* Kuo et al. [44]	2020	1 (Taiwan)	354	28	Convolutional neural network	I	TMS-4	Internal
	* Shi et al. [39]	2020	1 (China)	121	33	Neural network	P (49)	UHR-OCT & Pentacam HR	Internal
KC Severity	Yousefi et al. [49]	2018	multi-center (Japan)	3156		Density-based clustering	P (420)	CASIA OCT	NA

Study objectives: The aim of the research. It was either detecting KC from controls or detecting early KC from controls in this study. No. of Centers involved: The number of centers involved is reported, NR indicated the center is not reported explicitly. Data type (No. of parameters): The kind of data used as inputs to machine learning models. It was either images (graphics) or parameters in this study (numeric). The letter 'P' denoted parameters, while the letter 'I' denoted images. Corneal Imaging modality: Where the imaging system/systems of the input data was derived. Evaluation methods: Described how the model's performance was determined. External (evaluation in an independent database), internal (bootstrap validation, cross validation, random training test splits, temporal splits). Asterisks (\*) indicated studies that were excluded from the meta-analysis.



Of the 35 studies, 33 focused on developing machine learning models, while 2 [16,46] involved external validation of existing models (in a context other than that used for the model development). Four studies [19,22,31,37] developed and externally validated models in the same study (in a separate data set, eliminating random training test splits, and cross-validation), and used data from several centers. The remaining studies analyzed data from a single center (1 study [29] analyzed data from two centers) and internally validated the developed model (e.g., via cross validation, random training test splits). Studies were conducted in different countries, including the United States of America (USA) ( $n = 4$ ) [12,17,18,30], Japan ( $n = 3$ ) [19,42,49], France ( $n = 2$ ) [21,23], the United Kingdom (UK) ( $n = 1$ ) [24], India ( $n = 1$ ) [26], Brazil ( $n = 2$ ) [25,45], Belgium ( $n = 3$ ) [28,33,37], Hungary ( $n = 2$ ) [27,33], Austria ( $n = 1$ ) [41], Australia ( $n = 1$ ) (conducted by our group [48]), Spain ( $n = 3$ ) [35,36,40], Taiwan ( $n = 1$ ) [44], and China ( $n = 2$ ) [39,47], and 4 investigations [22,29,31,34] were cross-ethnicity studies.

### 3.3. Detecting KC from Controls and Meta-Analysis

A total of 26 studies [12,16–40] developed machine learning models that were based either directly on captured parameters or as calculated parameters from corneal topography or tomography systems. The number of parameters used in these studies ranged from 5 to 443 (Table 1). The machine learning algorithms explored in these studies included decision tree, discriminant analysis, logistic regression, naive bayes, neural networks, random forest, and support vector machine. The majority of these studies ( $n = 23$ ) employed a single machine learning algorithm, while 4 studies [24,29,31,38] compared several algorithms.

Four articles directly analyzed images [41–44] generated by corneal topography or tomography systems, as opposed to the image parameters. Convolutional neural networks, a common deep learning-based method, were used in all these studies.

Meta-analysis was performed on 26 of the 30 studies, with 4 studies [37,39,41,43] being excluded due to inadequate data needed to quantify the TP, FP, TN, and FN. Asterisks (\*) indicated studies that were excluded from the meta-analysis in Table 1. Deeks' funnel plot (Supplementary Figure S1) was used to assess possible publication bias. No evidence of publication bias was apparent ( $p = 0.91$ ).

The pooled sensitivity and specificity for KC versus control were 0.970 (95% CI 0.949–0.982) and 0.985 (95% CI 0.971–0.993). Pooled performance was computed using a bivariate random-effects model and represents a summary estimate of the sensitivity and specificity values (i.e., TP, FP, TN, and FN) obtained from each individual research [13]. The most frequently employed imaging technologies used were Pentacam (including Pentacam HR), TMS (including TMS-1 and TMS-4), and Orbscan (including Orbscan II and Orbscan IIz). Five studies analyzed data from the Pentacam, four studies from TMS, three studies from Orbscan, and other studies used data from Corvis, GALILEI, Sirius, or Keratron. For the studies based only on Pentacam data, the pooled sensitivity was 0.987 (95% CI 0.971–0.994) and the pooled specificity was 0.989 (95% CI 0.963–0.997). In the case of TMS, the pooled sensitivity was 0.943 (95% CI 0.897–0.969) and the pooled specificity was 0.978 (95% CI 0.954–0.989), whereas for the Orbscan data, the pooled sensitivity was 0.947 (95% CI 0.886–0.976) and the pooled specificity was 0.983 (95% CI 0.917–0.997) (Table 2).

**Table 2.** Diagnostic performance of artificial intelligence in detection of KC versus controls using different imaging modalities.

Imaging Modalities	Pooled Sensitivity	Pooled Specificity
Pentacam ( $n = 5$ )	0.987 (95% CI 0.971–0.994)	0.989 (95% CI 0.963–0.997)
TMS ( $n = 4$ )	0.943 (95% CI 0.897–0.969)	0.978 (95% CI 0.954–0.989)
Orbscan ( $n = 3$ )	0.947 (95% CI 0.886–0.976)	0.983 (95% CI 0.917–0.997)
Pooled total ( $n = 26$ )	0.970 (95% CI 0.949–0.982)	0.985 (95% CI 0.971–0.993)

### 3.4. Detecting Early KC from Controls and Meta-Analysis

A total of 13 studies [20,21,27–29,31,33,37,39,45–48] evaluated machine learning models to assess early KC from controls using corneal topography parameters. These studies used fewer parameters compared to those utilized to identify KC in controls, varying from 11 to 55 parameters (Table 1). Ten studies applied a single algorithm and three compared several algorithms (one of which was conducted by our group [48]). Additionally, one study [44] built the model by combining convolutional neural networks and TMS derived images.

Ten studies were included in the meta-analysis, out of a total of fourteen. As shown by an asterisk (\*) in Table 1, four studies [37,39,44,45] were omitted from the meta-analysis due to inadequate evidence to calculate the TP, FP, TN, and FN. Assessment of publishing bias with Deeks' funnel plot (Supplementary Figure S2) found no significant effect ( $p = 0.18$ ).

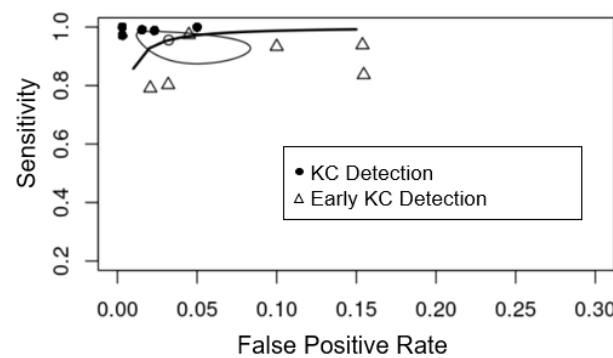
Overall, the pooled sensitivity was 0.882 (95% CI 0.822–0.923) and pooled specificity was 0.947 (95% CI 0.914–0.967) for early KC versus control. The most widely used imaging machines in the detection of early KC from control eyes were the Pentacam (including Pentacam HR) and Orbscan (including Orbscan II and Orbscan IIz). Six studies used Pentacam data, two studies used Orbscan data, and an additional two studies used data from either GALILEI, or a combination of Pentacam and Corvis. Early KC detection (Table 3) was correlated with a pooled sensitivity of 0.882 (95% CI 0.795–0.935), and a pooled specificity of 0.935 (95% CI 0.874–0.967) for models utilizing Pentacam data. When using Orbscan data, a pooled sensitivity of 0.842 (95% CI 0.504–0.965), and a pooled specificity of 0.958 (95% CI 0.821–0.991) were obtained.

**Table 3.** Diagnostic performance of machine learning on detection early KC using different imaging modalities.

Imaging Modalities	Pooled Sensitivity	Pooled Specificity
Pentacam ( $n = 6$ )	0.882 (95% CI 0.795–0.935)	0.935 (95% CI 0.874–0.967)
Orbscan ( $n = 2$ )	0.842 (95% CI 0.504–0.965)	0.958 (95% CI 0.821–0.991)
Pooled total ( $n = 10$ )	0.882 (95% CI 0.822–0.923)	0.947 (95% CI 0.914–0.967)

The pooled diagnostic performance of detecting KC compared to controls was superior to that of early KC with sensitivity (0.970 (95% CI 0.949–0.982) vs. 0.882 (95% CI 0.822–0.923) and specificity (0.985 (95% CI 0.971–0.993) vs. 0.947 (95% CI 0.914–0.967)). This difference implied that early KC detection using machine learning algorithms are still in their infancy.

The Pentacam tomography system was the most commonly used corneal imaging device for both the KC and early KC categories. In Figure 4, we compare studies that used the Pentacam data set to diagnose KC and early KC. Diagnostic efficiency for detecting KC was superior to that for early KC, as shown by the higher sensitivity (y-axis) and specificity (x-axis) in the plot.



**Figure 4.** Summary receiver-operating characteristic curves of the diagnostic performance of machine learning detecting KC (black circle) and early KC (triangle) from controls using Pentacam parameters. The white circle is the summary estimate point (sensitivity (0.956 (95% CI 0.897–0.982), specificity (0.968 (95% CI 0.931–0.985)) of studies using Pentacam parameters. The Y-axis represents sensitivity, with higher values indicating greater sensitivity, and the X-axis represents false positive rate, which was equal to 1-specificity, with lower values indicating greater specificity.

### 3.5. Detection of Different KC Severities

Along with detecting KC eyes as a distinct category, six of the published studies grouped their KC eyes into clinical stages and used machine learning algorithms to identify each stage separately. These studies classified KC eyes into various categories based on a variety of measures. Kamiya et al. [42] classified eyes into Grades 1–4 according to the Amsler–Krumeich (AK) classification scheme, which is mostly focused on keratometry, but often incorporates refraction and pachymetry [51]. Bolarin et al. [35] and Velazquez-Blazquez et al. [40] graded eyes into Grade I–IV plus or mild KC, using a different classification system named RETICS, based on corrected distance visual acuity (CDVA) [51]. Another study, Issarti et al. [33] classified their KC eyes into mild and moderate stages using a self-defined classification (described at the end of Table 4), whereas Maeda et al. [12] did not specify their staging method. Table 4 summarizes these findings. Since no consistent grading system was used for classifying KC severity in these studies—and indeed none is globally established [52]—the findings were therefore not directly comparable.

**Table 4.** Characteristics of machine learning-assisted studies for detection of KC severity.

First Author	Year	Severity Grading (No. of Eyes)	Definition/Classification Methods	Corneal Imaging Modality	Reported Sensitivity in Detection of Each Severity Level
Maeda et al. [12]	1995	Mild (15) Moderate (18) Advanced (11)	NA	TMS-1	Mild: 100% Moderate: 100% Advanced: 91%
Kamiya et al. [42]	2019	Grade 1 (108) Grade 2 (75) Grade 3 (42) Grade 4 (79)	Amsler–Krumeich classification	AS-OCT	Grade 1: 88.9% Grade 2: 68% Grade 3: 71.4% Grade 4: 74.7%
Issarti et al. [33]	2019	Mild KC (220)	<sup>a</sup> Self-defined	Pentacam	98.81%
Issarti et al. [33]	2019	Moderate KC (229)	<sup>b</sup> Self-defined	Pentacam	99.91%
Bolarin et al. [35]	2020	Grade I (44) Grade II (18) Grade III (15) Grade IV (15) Grade IV plus (15)	RETICS grading	Sirius	Grade I: 59.1% Grade II: 33.3% Grade III: 40% Grade IV: 80% Grade IV plus: 86.7%
Velazquez-Blazquez et al. [40]	2020	Mild KC (42)	RETICS grading	Sirius	Mild KC: 63%

<sup>a</sup> A clear cornea, tomography maps compatible with KC, a Fleischer ring at the apex base, slight thinning, and anterior and/or posterior corneal steepening; <sup>b</sup> Slit-lamp findings compatible with KC, corneal thinning at the apex, Vogt striae, a clearly visible Fleischer ring and corneal tomography compatible with KC; The severity of KC was considered to be increasing from Grade 1 to Grade 4 and for Grade I to Grade IV plus.

The first attempt at proposing a data-driven KC classification scheme was based on density-based clustering by Yousefi et al. [49] using OCT-based parameters from 3156 eyes. They identified five clusters that could represent five stages of KC, ranging from normal to advanced KC.

### 3.6. Reporting Completeness of Machine Learning Studies in KC

Publications had only low to moderate adherence to TRIPOD items, ranging between 3% and 48%, with a median of 28%. Each item on the TRIPOD checklist was adhered to by between 0% and 69% of studies, with an average (median) of 23% (Figure 5). Eight items were reported in over 50% of studies, whereas ten items were reported in fewer than 10% of studies. Supplementary Table S1 details results of each TRIPOD item for each paper and the level of reporting adherence for each TRIPOD item. The title, abstract, predictor assessment, management of missing data, model performance assessment, and description of participant characteristics were the most poorly rated items, with no research fulfilling all requirements. In 69% of studies (24/35) [16,17,21–24,27,30–33,35–40,42–44,46–49], the medical context explanation was well stated, elucidating the reason for creating the models for detecting KC. Moreover, in 66% (23/35) of studies [12,16,17,20–23,25,28,30–34,36,41–44,46–49], the type of transformation of continuous predictors (e.g., linear or nonlinear) were specified.

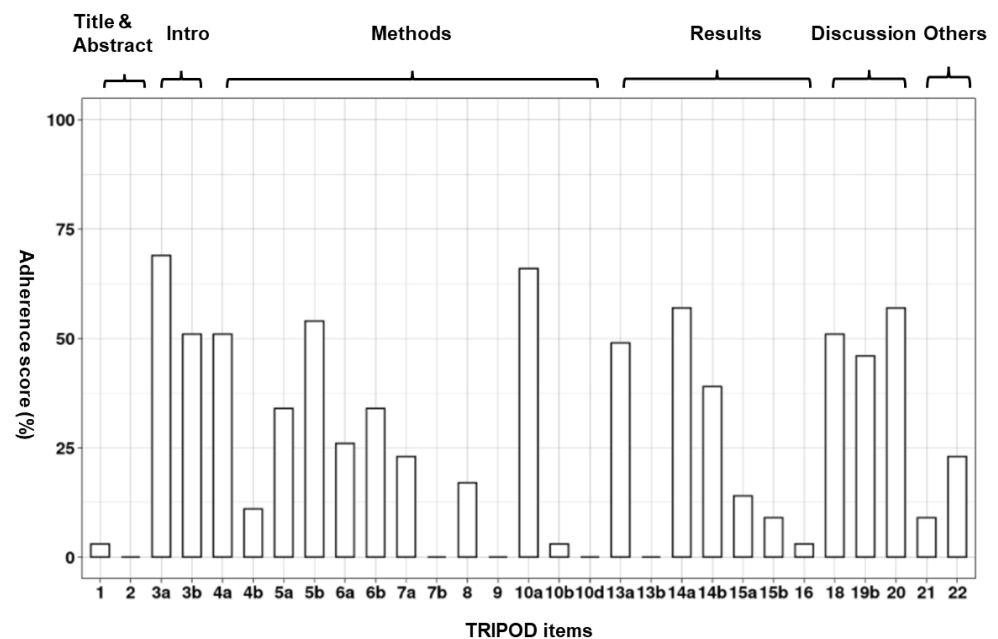


Figure 5. Overall adherence per TRIPOD item.

## 4. Discussion

This is the first comprehensive meta-analysis on KC and machine learning, and it has demonstrated that by using data from a variety of corneal imaging devices, machine learning can reliably distinguish KC eyes from control eyes (pooled sensitivity > 0.90). However, the performance of machine learning models in distinguishing early KC eyes from controls was poorer, with the maximum pooled sensitivity of 0.88. Overall completeness of model was evaluated using the TRIPOD guidelines, and standard reporting compliance was found to be inadequate in all published KC machine learning research so far undertaken.

Machine learning models in KC were developed for a variety of imaging systems. Models themselves are therefore not directly interchangeable owing to the different input expectations. This may have a detrimental effect on the clinical translation of these models. For example, Smadja et al. developed a machine learning model that had a sensitivity of 93.6% and a specificity of 97.2% for discriminating normal eyes from early KC in their study [21]. This model was constructed using the anterior and posterior asphericity asymmetry indices (AAI and OSI), corneal volume, paracentral mean keratometry, and anterior

chamber depth derived from a GALILEI machine. Parameters such as AAI and OSI are not accessible in other corneal tomography imaging systems, such as Pentacam [53], suggesting that the generated model cannot be utilized in clinics equipped with the Pentacam system. Pentacam is one of the most frequently used corneal tomographic technologies in clinical practice [11], and our research discovered that the Pentacam is the most frequently utilized source of data in the publications we reviewed [26–28,31,33,37,47,48]. As a consequence, despite the fact that the model created by Smadja et al. showed a high level of performance for early KC diagnosis, it is not generally applicable.

The machine learning models that used data from the Pentacam demonstrated a higher pooled sensitivity and specificity in detecting KC and early KC from control eyes compared to other imaging machines. This is likely due to the ability of the Pentacam machine to generate a wider spectrum of data than other systems, including data on the front cornea, the back cornea, corneal pachymetry, and other areas of the anterior eye segment [54,55].

A review of the literature on the application of AI to evaluate corneal topography for the diagnosis and early detection of corneal ectasias was recently published [56]. That article summarized significant advances in corneal imaging and the application of AI in KC as viewed by an Eye Care Professional, a biomedical engineer, and a data scientist. It concluded that AI in corneal imaging may improve refractive surgery and diagnosis of corneal ectasias. That review focused exclusively on corneal imaging modalities and their performance in relation to AI. In contrast to that study, our systematic review included meta-analysis-additional information on individual studies as well as the pooled performance of existing machine learning models for detecting KC as well as early KC. In addition, we also reported on the assessment of completeness of items included in the machine learning models using the TRIPOD reporting system. We analyzed each study by its objectives and showed that while substantial research had been undertaken on KC detection, much less had been undertaken on early KC detection, KC severity detection, and detection of progression. Our analysis also evaluated the parameters utilized in prior studies, indicating that no study had analyzed all available parameters despite the extensive use of corneal imaging data. Additionally, we compared and contrasted the corneal imaging systems used. These aspects have not been previously reviewed.

#### *4.1. Reporting Completeness of Machine Learning Studies in KC*

Only when all model components are completely and transparently reported can the model's potential clinical usefulness be appropriately assessed. The main objective of machine learning models is to help clinicians in making medical decisions about an individual patient [57]. Users or doctors will need information on the clinical setting in which the diagnosis is needed (e.g., primary care, secondary care, or the general population), as well as the patients for whom the model is appropriate. Additionally, they will need information about which clinical data, referred to as predictors in the model, are necessary for model usage, as well as the definition of the patient outcome to which the model is referring. Unfortunately, this systematic review suggests that the studies on machine learning and KC often lacked sufficient description. Only 34% provided target setting [12,19,24,30,31,33,35,39,42,44,47,49], 23% defined all necessary predictors precisely [17,21,23,25,34,39,45,48], 26% exactly defined the patient outcome [21,28,29,39,44,46–49], and none completely described participant characteristics (e.g., basic demographics, clinical features, and available predictors).

Attempting to replicate the process of model construction using one's own data, machine learning researchers working in KC will need to specify information on the approach used to handle any missing data, the full prediction model, and a comprehensive explanation of all model-building procedures. However, only one study (3%) defined the whole model-building process [28], 14% provided the complete prediction model [20,21,26,35,47], and none explained how missing data were handled explicitly.

One factor contributing to the low percentage of adherence is that not all elements are relevant to all machine learning research. For example, it is impractical to mathematically

specify the final model in a study using non-regression techniques such as random forest, support vector machine, or neural network. While this is a legitimate rationale, it is important to refer to relevant TRIPOD items and provide as much information about model creation as possible in order to guide the modelling choices and facilitate subsequent validation.

#### 4.2. Limitations in the Current Literature

For the studies that reported the number of parameters employed, the range was between 5 and 443. Ruiz et al. used 22 Pentacam parameters to construct a machine learning model for identifying KC and early KC [28]. The parameters they employed, such as corneal curvature and pachymetry, were mostly those found on Pentacam's four-map selectable report, which is widely used in clinics to assess KC. As noted in their research, these 22 clinically relevant parameters were chosen from a pool of over 1000 accessible from Pentacam. Such pre-modelling parameter selection was observed in the majority of reviewed studies [12,17–24,26,27,29–40,42–45,47–49]; although, it was unclear why certain predictors were chosen as inputs. As a consequence of this filtering, the role of unexplored parameters in the identification of KC has remained unclear.

While the majority of research developed novel models for KC detection, only four [19,22,31,37] conducted external validation using other data sources. One common concern is that local data sets used for validation are unlikely to be representative of the target population on a global scale [57]. When evaluated on data sets collected in the United States of America and Switzerland, the machine learning model developed in a cross-ethnic research by Mahmoud et al. [22] performed differently. This implies that most of the current identified models cannot be used in a broad clinical setting since their performance may vary and it will most likely be poorer when applied to other external clinics or nations. We recommend that any model be externally evaluated on a large scale to understand this variation. This is especially true given the relatively small sample sizes reported in each study; international collaboration would therefore be highly advantageous to move this field along.

The development of KC is frequently manifested not only by corneal changes, but also by clinical symptoms such as vision, refraction, and slit-lamp findings. No study has ever incorporated all pertinent data [12,16–49]. Demographic data, such as age and gender, as well as potential risk factors for KC, such as eye rubbing and family history, may also aid in KC detection. We have limited knowledge of how these critical factors may influence the detection of KC in the machine learning models currently employed. Due to the lack of known risk factors for KC [2], it should be possible to start assessing potential risk factors and include in future machine learning models to assess their importance in diagnosis.

The bulk of reported studies of KC have used data derived from a single corneal topography or tomography imaging device to train their machine learning models. Thus, there is a dearth of information as to what impact combining data from multiple devices would have on machine learning models in the detection of KC. Other forms of data, such as from the corneal epithelial thickness map produced with optical coherence tomography (OCT) and corneal biomechanical measurements, are also increasingly being recognized as crucial in the diagnosis of KC [30], particularly early KC [58]. Integrating data from multiple devices and considering a broader variety of factors may therefore further improve the early detection of KC.

The majority of studies employed a specific machine learning method; although, there was limited information on how or why the authors chose a particular algorithm from a large number of potential alternatives. Given that machine learning algorithms are sometimes referred to as 'black boxes', it would be advantageous to evaluate several machine learning algorithms on the same data set and select the best one [59]. This process may assist in avoiding selection bias and may aid in improving early KC detection.

Imbalanced sample sizes were found in about half of the studies analyzed [12,16,20,21,27–31,33,44,46], which could skew the machine learning model and

impair its capacity to identify cases. In comparison to the number of control eyes, most studies included a lower number of KC or early KC eyes for model establishment. This may be a concern when considering the performance of the model. For example, in the study by Lopes et al. [31], a random forest model was constructed using 2980 stable (control) eyes and 71 ectatic susceptibility eyes, achieving a sensitivity of 80% and a specificity of 96.8%. However, their model's baseline accuracy, defined as the ability to identify all eyes in the control group without constructing a model, was 97.7%. Although the analysis obtained a high specificity of 96.8%, this could be deemed as suboptimal in comparison to its potential baseline accuracy. Instead of utilizing accuracy, sensitivity, and specificity, precision value (also known as positive predictive value) may provide a more interpretable evaluation for models with unbalanced sample sizes. This metric has been employed in several studies to evaluate their models [22,28,32,33,36,47,48].

There have been fewer studies on identifying the different stages of KC, and a standardized classification system for KC has yet to be devised. There has been no machine learning study using longitudinal data on KC progression. While applications of machine learning have benefited in predicting and detecting progression indicators in AMD, diabetic retinopathy, and glaucoma [60–64]. The combination of machine learning methods and large clinical data sets may assist in the analysis of KC progression.

#### 4.3. Approach for Future Studies

Machine learning has been increasingly used in KC over the last three decades, mostly for the identification of KC and early KC. The advantage of machine learning is that it allows consistent and unbiased diagnosis, which is critical when diagnosing patients at an early stage, as early intervention using treatments such as corneal crosslinking (CXL), could delay or slow disease progression, thus preventing the need for a possible corneal transplant.

There is still room to improve the efficiency of machine learning models in detecting early KC. This may be accomplished by allowing the use of all publicly available data, including complete databases from corneal imaging systems, clinical data, genetic data, and other risk factors. There is also a need to maximize the potential of machine learning techniques by optimizing their output at the methodology and data space levels.

Currently, there are no successful examples of machine learning models being used in clinical practice. This may be due to a lack of large patient populations to validate results, the utilization of various imaging devices, a local participant group comprised of individuals of various ethnic backgrounds, clinicians' overall acceptance of machine learning techniques for diagnosis and their relative reliability to humans. External model validation on a diverse patient population, as well as the creation of platform-independent models that can be generalized through several corneal imaging systems, are therefore needed. Finally, research on machine learning in KC should also address additional research gaps in the area, such as classifying KC severity, and identifying and forecasting KC progression.

## 5. Conclusions

We present an up to date, comprehensive review on the use of machine learning in KC detection and identify the substantial limitations that need to be overcome to make the diagnostic process more efficient for early keratoconus. In light of our findings on pooling detection performance and low adherence to the TRIPOD checklist, we believe that both improved machine learning model performance in early KC detection and improved quality machine learning research in KC is sorely needed. Despite various challenges, the future of integrating machine learning technologies into clinical practice is promising with the advent of advanced imaging modalities. Machine learning can be further investigated for broad application to the entire process of KC detection and management. In particular, some open avenues for research include early KC detection, risk factor evaluation, prediction of progression, and clinical management guidance. However, global collaboration is essential to obtain larger data sets and more robust models.

**Supplementary Materials:** The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/jcm11030478/s1>, Figure S1: Assessment of publication risk bias across studies using Deeks' funnel plot asymmetry test in detecting KC from controls. ESS, effective sample size, which is determined as a function of the number of diseased ( $n_1$ ) and non-diseased ( $n_2$ ) subjects:  $(4n_1 * n_2)/(n_1 + n_2)$ . The y-axis shows the inverse of the square root of the effective sample size ( $1/\text{root}(\text{ESS})$ ). The x-axis shows the natural logarithm of the diagnostic odds ratio ( $\ln\text{DOR} (\text{TP} * \text{TN})/(\text{FP} * \text{FN})$ ). In this figure, the vertical line represents the meta-analysis summary estimate, and each black circle represents one study, and dispersed symmetrically right and left the vertical line implies a low probability of publication bias for the included research. The dotted line indicates the ESS weighted regression tests of funnel plot asymmetry, and  $p$  values  $< 0.05$  were considered as significant; Figure S2: Assessment of publication risk bias across studies using Deeks' funnel plot asymmetry test in detecting early KC from controls. ESS, effective sample size, which is determined as a function of the number of diseased ( $n_1$ ) and non-diseased ( $n_2$ ) subjects:  $(4n_1 * n_2)/(n_1 + n_2)$ . The y-axis shows the inverse of the square root of the effective sample size ( $1/\text{root}(\text{ESS})$ ). The x-axis shows the natural logarithm of the diagnostic odds ratio ( $\ln\text{DOR} (\text{TP} * \text{TN})/(\text{FP} * \text{FN})$ ). In this figure, the vertical line represents the meta-analysis summary estimate, and each black circle represents one study, and dispersed symmetrically right and left the vertical line implies a low probability of publication bias for the included research. The dotted line indicates the ESS weighted regression tests of funnel plot asymmetry, and  $p$  values  $< 0.05$  were considered as significant; Table S1: Results of each TRIPOD item for each paper and the level of reporting adherence for each TRIPOD item.

**Author Contributions:** Conceptualization, P.N.B. and S.S.; methodology, K.C.; software, K.C.; formal analysis, K.C.; data curation, K.C. and S.S.; writing—original draft preparation, K.C.; writing—review and editing, P.N.B., S.S., K.V. and K.C.; supervision, P.N.B., S.S. and K.V.; funding acquisition, P.N.B. and S.S. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the Australian National Health and Medical Research Council (NHMRC) project Ideas grant, grant number APP1187763, Senior Research Fellowship grant number 1138585 to PNB, Lions Eye Donation Service (SS), Angior Family Foundation (SS), Perpetual Impact Philanthropy grant (SS), and Keratoconus Australia Funding (SS). The Centre for Eye Research Australia (CERA) receives Operational Infrastructure Support from the Victorian Government. The sponsor or funding organizations had no role in the design or conduct of this research.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** The corresponding author had full access to all the data in the study; data are available upon reasonable request.

**Acknowledgments:** The authors would like to thank Myra McGuinness from the Centre for Eye Research Australia for her assistance with the protocol registration.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Whitcher, J.P.; Srinivasan, M.; Upadhyay, M.P. Corneal blindness: A global perspective. *Bull. World Health Organ.* **2001**, *79*, 214–221. [[PubMed](#)]
2. Sharif, R.; Bak-Nielsen, S.; Hjortdal, J.; Karamichos, D. Pathogenesis of Keratoconus: The intriguing therapeutic potential of Prolactin-inducible protein. *Prog. Retin. Eye Res.* **2018**, *67*, 150–167. [[CrossRef](#)] [[PubMed](#)]
3. Buzzonetti, L.; Bohringer, D.; Liskova, P.; Lang, S.; Valente, P. Keratoconus in Children: A Literature Review. *Cornea* **2020**, *39*, 1592–1598. [[CrossRef](#)] [[PubMed](#)]
4. Röck, T.; Bartz-Schmidt, K.U.; Röck, D. Trends in corneal transplantation at the University Eye Hospital in Tübingen, Germany over the last 12 years: 2004–2015. *PLoS ONE* **2018**, *13*, e0198793. [[CrossRef](#)]
5. Kelly, T.L.; Williams, K.A.; Coster, D.J.; Australian Corneal Graft Registry. Corneal transplantation for keratoconus: A registry study. *Arch. Ophthalmol.* **2011**, *129*, 691–697. [[CrossRef](#)]
6. Andreu-Perez, J.; Poon, C.C.; Merrifield, R.D.; Wong, S.T.; Yang, G.Z. Big data for health. *IEEE J. Biomed. Health Inform.* **2015**, *19*, 1193–1208. [[CrossRef](#)]
7. Verspoor, K.; Martin-Sanchez, F. Big Data in Medicine is Driving Big Changes. *Yearb. Med. Inform.* **2014**, *23*, 14–20. [[CrossRef](#)]
8. Mintz, Y.; Brodie, R. Introduction to artificial intelligence in medicine. *Minim. Invasive Ther. Allied Technol.* **2019**, *28*, 73–81. [[CrossRef](#)]
9. Rajkomar, A.; Dean, J.; Kohane, I. Machine learning in medicine. *N. Engl. J. Med.* **2019**, *380*, 1347–1358. [[CrossRef](#)]



10. Ramamurthy, S.; Reddy, J.; Jhanji, V. Topography and tomography in the diagnosis of corneal ectasia. *Expert Rev. Ophthalmol.* **2015**, *10*, 215–228. [[CrossRef](#)]
11. Motlagh, M.N.; Moshirfar, M.; Murri, M.S.; Skanchy, D.F.; Momeni-Moghaddam, H.; Ronquillo, Y.C.; Hoopes, P.C. Pentacam® Corneal Tomography for Screening of Refractive Surgery Candidates: A Review of the Literature, Part I. *Med. Hypothesis Discov. Innov. Ophthalmol.* **2019**, *8*, 177–203. [[PubMed](#)]
12. Maeda, N.; Klyce, S.D.; Smolek, M.K. Comparison of Methods for Detecting Keratoconus Using Videokeratography. *Arch. Ophthalmol.* **1995**, *113*, 870–874. [[CrossRef](#)] [[PubMed](#)]
13. Reitsma, J.B.; Glas, A.S.; Rutjes, A.W.S.; Scholten, R.J.; Bossuyt, P.M.; Zwinderman, A.H. Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. *J. Clin. Epidemiol.* **2005**, *58*, 982–990. [[CrossRef](#)]
14. van Enst, W.A.; Ochodo, E.; Scholten, R.J.; Hooft, L.; Leeftang, M.M. Investigation of publication bias in meta-analyses of diagnostic test accuracy: A meta-epidemiological study. *BMC Med. Res. Methodol.* **2014**, *14*, 70. [[CrossRef](#)]
15. Deeks, J.J.; Macaskill, P.; Irwig, L. The performance of tests of publication bias and other sample size effects in systematic reviews of diagnostic test accuracy was assessed. *J. Clin. Epidemiol.* **2005**, *58*, 882–893. [[CrossRef](#)] [[PubMed](#)]
16. Kalin, N.S.; Maeda, N.; Klyce, S.D.; Hargrave, S.; Wilson, S.E. Automated topographic screening for keratoconus in refractive surgery candidates. *CLAO J. Off. Publ. Contact Lens Assoc. Ophthalmol. Inc.* **1996**, *22*, 164–167.
17. Rabinowitz, Y.S.; Rasheed, K.; Yang, H.; Elashoff, J. Accuracy of ultrasonic pachymetry and videokeratography in detecting keratoconus. *J. Cataract. Refract. Surg.* **1998**, *24*, 196–201. [[CrossRef](#)]
18. Twa, M.D.; Parthasarathy, S.; Roberts, C.; Mahmoud, A.M.; Raasch, T.W.; Bullimore, M.A. Automated decision tree classification of corneal shape. *Optom. Vis. Sci. Off. Publ. Am. Acad. Optom.* **2005**, *82*, 1038–1046. [[CrossRef](#)]
19. Bessho, K.; Maeda, N.; Kuroda, T.; Fujikado, T.; Tano, Y.; Oshika, T. Automated keratoconus detection using height data of anterior and posterior corneal surfaces. *Jpn. J. Ophthalmol.* **2006**, *50*, 409–416. [[CrossRef](#)] [[PubMed](#)]
20. Saad, A.; Gatinel, D. Topographic and Tomographic Properties of Forme Fruste Keratoconus Corneas. *Investig. Ophthalmol. Vis. Sci.* **2010**, *51*, 5546–5555. [[CrossRef](#)]
21. Smadja, D.; Touboul, D.; Cohen, A.; Doveh, E.; Santhiago, M.R.; Mello, G.R.; Krueger, R.R.; Colin, J. Detection of Subclinical Keratoconus Using an Automated Decision Tree Classification. *Am. J. Ophthalmol.* **2013**, *156*, 237–246.e1. [[CrossRef](#)]
22. Mahmoud, A.M.; Nuñez, M.X.; Blanco, C.; Koch, D.D.; Wang, L.; Weikert, M.; Frueh, B.E.; Tappeiner, C.; Twa, M.; Roberts, C.J. Expanding the Cone Location and Magnitude Index to Include Corneal Thickness and Posterior Surface Information for the Detection of Keratoconus. *Am. J. Ophthalmol.* **2013**, *156*, 1102–1111. [[CrossRef](#)]
23. Saad, A.; Guilbert, E.; Gatinel, D. Corneal Enantiomorphism in Normal and Keratoconic Eyes. *J. Refract. Surg.* **2014**, *30*, 542–547. [[CrossRef](#)]
24. Silverman, R.H.; Urs, R.; RoyChoudhury, A.; Archer, T.J.; Gobbe, M.; Reinstein, D.Z. Epithelial Remodeling as Basis for Machine-Based Identification of Keratoconus. *Investig. Ophthalmol. Vis. Sci.* **2014**, *55*, 1580–1587. [[CrossRef](#)] [[PubMed](#)]
25. Koprowski, R.; Ambrosio, R. Quantitative assessment of corneal vibrations during intraocular pressure measurement with the air-puff method in patients with keratoconus. *Comput. Biol. Med.* **2015**, *66*, 170–178. [[CrossRef](#)]
26. Shetty, R.; Matalia, H.; Srivatsa, P.; Ghosh, A.; Dupps, W.; Roy, A.S. A Novel Zernike Application to Differentiate between Three-dimensional Corneal Thickness of Normal Corneas and Corneas with Keratoconus. *Am. J. Ophthalmol.* **2015**, *160*, 453–462.e2. [[CrossRef](#)]
27. Kovács, I.; Miháltz, K.; Kránitz, K.; Juhász, É.; Takács, Á.; Dienes, L.; Gergely, R.; Nagy, Z.Z. Accuracy of machine learning classifiers using bilateral data from a Scheimpflug camera for identifying eyes with preclinical signs of keratoconus. *J. Cataract. Refract. Surg.* **2016**, *42*, 275–283. [[CrossRef](#)]
28. Hidalgo, I.R.; Rodriguez, P.; Rozema, J.J.; Dhubhghaill, S.N.; Zakaria, N.; Tassignon, M.J.; Koppen, C. Evaluation of a Machine-Learning Classifier for Keratoconus Detection Based on Scheimpflug Tomography. *Cornea* **2016**, *35*, 827–832. [[CrossRef](#)]
29. Ambrósio, R.; Lopes, B.T.; Faria-Correia, F.; Salomão, M.Q.; Bühren, J.; Roberts, C.J.; Elsheikh, A.; Vinciguerra, R.; Vinciguerra, P. Integration of Scheimpflug-Based Corneal Tomography and Biomechanical Assessments for Enhancing Ectasia Detection. *J. Refract. Surg.* **2017**, *33*, 434–443. [[CrossRef](#)]
30. Silverman, R.H.; Urs, R.; Roychoudhury, A.; Archer, T.J.; Gobbe, M.; Reinstein, D.Z. Combined tomography and epithelial thickness mapping for diagnosis of keratoconus. *Eur. J. Ophthalmol.* **2017**, *27*, 129–134. [[CrossRef](#)] [[PubMed](#)]
31. Lopes, B.T.; Ramos, I.C.; Salomão, M.Q.; Guerra, F.P.; Schallhorn, S.C.; Schallhorn, J.M.; Vinciguerra, R.; Vinciguerra, P.; Price, F.W., Jr.; Price, M.O.; et al. Enhanced Tomographic Assessment to Detect Corneal Ectasia Based on Artificial Intelligence. *Am. J. Ophthalmol.* **2018**, *195*, 223–232. [[CrossRef](#)]
32. Chandapura, R.; Salomao, M.Q.; Ambrosio, R.; Swarup, R.; Shetty, R.; Roy, A.S. Bowman’s topography for improved detection of early ectasia. *J. Biophotonics* **2019**, *12*, e201900126. [[CrossRef](#)] [[PubMed](#)]
33. Issarti, I.; Consejo, A.; Jiménez-García, M.; Hershko, S.; Koppen, C.; Rozema, J. Computer aided diagnosis for suspect keratoconus detection. *Comput. Biol. Med.* **2019**, *109*, 33–42. [[CrossRef](#)] [[PubMed](#)]
34. Leão, E.; Ren, T.I.; Lyra, J.M.; Machado, A.; Koprowski, R.; Lopes, B.; Vinciguerra, R.; Vinciguerra, P.; Roberts, C.J.; Elsheikh, A.; et al. Corneal deformation amplitude analysis for keratoconus detection through compensation for intraocular pressure and integration with horizontal thickness profile. *Comput. Biol. Med.* **2019**, *109*, 263–271. [[CrossRef](#)] [[PubMed](#)]
35. Bolarín, J.M.; Cavas, F.; Velázquez, J.; Alió, J. A Machine-Learning Model Based on Morphogeometric Parameters for RETICS Disease Classification and GUI Development. *Appl. Sci.* **2020**, *10*, 1874. [[CrossRef](#)]

36. Castro-Luna, G.M.; Martínez-Finkelshtein, A.; Ramos-López, D. Robust keratoconus detection with Bayesian network classifier for Placido-based corneal indices. *Contact Lens Anterior Eye* **2020**, *43*, 366–372. [CrossRef]
37. Issarti, I.; Consejo, A.; Jiménez-García, M.; Kreps, E.O.; Koppen, C.; Rozema, J.J. Logistic index for keratoconus detection and severity scoring (Logik). *Comput. Biol. Med.* **2020**, *122*, 103809. [CrossRef] [PubMed]
38. Lavric, A.; Popa, V.; Takahashi, H.; Yousefi, S. Detecting Keratoconus from Corneal Imaging Data Using Machine Learning. *IEEE Access* **2020**, *8*, 149113–149121. [CrossRef]
39. Shi, C.; Wang, M.; Zhu, T.; Zhang, Y.; Ye, Y.; Jiang, J.; Chen, S.; Lu, F.; Shen, M. Machine learning helps improve diagnostic ability of subclinical keratoconus using Scheimpflug and OCT imaging modalities. *Eye Vis.* **2020**, *7*, 48. [CrossRef]
40. Velázquez-Blázquez, J.S.; Bolarín, J.M.; Cavas-Martínez, F.; Alió, J.L. EMKLAS: A New Automatic Scoring System for Early and Mild Keratoconus Detection. *Transl. Vis. Sci. Technol.* **2020**, *9*, 30. [CrossRef]
41. Dos Santos, V.A.; Schmetterer, L.; Stegmann, H.; Pfister, M.; Messner, A.; Schmidinger, G.; Garhofer, G.; Werkmeister, R.M. CorneaNet: Fast segmentation of cornea OCT scans of healthy and keratoconic eyes using deep learning. *Biomed. Opt. Express* **2019**, *10*, 622–641. [CrossRef]
42. Kamiya, K.; Ayatsuka, Y.; Kato, Y.; Fujimura, F.; Takahashi, M.; Shoji, N.; Mori, Y.; Miyata, K. Keratoconus detection using deep learning of colour-coded maps with anterior segment optical coherence tomography: A diagnostic accuracy study. *BMJ Open* **2019**, *9*, e031313. [CrossRef]
43. Lavric, A.; Valentin, P. KeratoDetect: Keratoconus Detection Algorithm Using Convolutional Neural Networks. *Comput. Intell. Neurosci.* **2019**, *2019*, 8162567. [CrossRef]
44. Kuo, B.-I.; Chang, W.-Y.; Liao, T.-S.; Liu, F.-Y.; Liu, H.-Y.; Chu, H.-S.; Chen, W.-L.; Hu, F.-R.; Yen, J.-Y.; Wang, I.-J. Keratoconus Screening Based on Deep Learning Approach of Corneal Topography. *Transl. Vis. Sci. Technol.* **2020**, *9*, 53. [CrossRef] [PubMed]
45. Ventura, B.V.; Machado, A.P.; Ambrósio, R.; Ribeiro, G.; Araújo, L.N.; Luz, A.; Lyra, J.M. Analysis of Waveform-Derived ORA Parameters in Early Forms of Keratoconus and Normal Corneas. *J. Refract. Surg.* **2013**, *29*, 637–643. [CrossRef]
46. Chan, C.; Ang, M.; Saad, A.; Chua, D.; Mejia, M.; Lim, L.; Gatinel, D. Validation of an Objective Scoring System for Forme Fruste Keratoconus Detection and Post-LASIK Ectasia Risk Assessment in Asian Eyes. *Cornea* **2015**, *34*, 996–1004. [CrossRef] [PubMed]
47. Xu, Z.; Li, W.; Jiang, J.; Zhuang, X.; Chen, W.; Peng, M.; Wang, J.; Lu, F.; Shen, M.; Wang, Y. Characteristic of entire corneal topography and tomography for the detection of sub-clinical keratoconus with Zernike polynomials using Pentacam. *Sci. Rep.* **2017**, *7*, 16486. [CrossRef] [PubMed]
48. Cao, K.; Verspoor, K.; Sahebjada, S.; Baird, P.N. Evaluating the Performance of Various Machine Learning Algorithms to Detect Subclinical Keratoconus. *Transl. Vis. Sci. Technol.* **2020**, *9*, 24. [CrossRef] [PubMed]
49. Yousefi, S.; Yousefi, E.; Takahashi, H.; Hayashi, T.; Tampo, H.; Inoda, S.; Arai, Y.; Asbell, P. Keratoconus severity identification using unsupervised machine learning. *PLoS ONE* **2018**, *13*, e0205998. [CrossRef]
50. Henriquez, M.A.; Hadid, M.; Izquierdo, L., Jr. A systematic review of subclinical keratoconus and forme fruste keratoconus. *J. Refract. Surg.* **2020**, *36*, 270–279. [CrossRef]
51. Alió, J.L.; Vega-Estrada, A.; Sanz-Diez, P.; Peña-García, P.; Durán-García, M.L.; Maldonado, M. Keratoconus management guidelines. *Int. J. Keratoconus Ectatic Corneal Dis.* **2015**, *4*, 1–39.
52. Gomes, J.A.; Tan, D.; Rapuano, C.J.; Belin, M.W.; Ambrósio, R., Jr.; Guell, J.L.; Malecaze, F.; Nishida, K.; Sangwan, V.S. Global consensus on keratoconus and ectatic diseases. *Corneal* **2015**, *34*, 359–369. [CrossRef]
53. Pentacam User Guide. System for Measuring and Analysing the Front Part of the Eye. User Guide Pentacam®/Pentacam®HR/Pentacam®AXL. Available online: [https://www.pentacam.com/fileadmin/user\\_upload/pentacam.de/downloads/interpretations-leitfaden/interpretation\\_guideline\\_3rd\\_edition\\_0915.pdf](https://www.pentacam.com/fileadmin/user_upload/pentacam.de/downloads/interpretations-leitfaden/interpretation_guideline_3rd_edition_0915.pdf)2018 (accessed on 8 June 2021).
54. Chen, D.; Lam, A.K. Reliability and repeatability of the Pentacam on corneal curvatures. *Clin. Exp. Optom.* **2009**, *92*, 110–118. [CrossRef]
55. Michael, W.; Belin, M. Point/Counterpoint: The Pentacam Versus the Orbscan. 2006. Available online: <https://crstoday.com/> (accessed on 8 June 2021).
56. Shanthi, S.; Aruljyothi, L.; Balasundaram, M.B.; Janakiraman, A.; Nirmala, D.K.; Pyingkodi, M. Artificial intelligence applications in different imaging modalities for corneal topography. *Surv. Ophthalmol.* **2021**. [CrossRef]
57. Wynants, L.; Van Calster, B.; Collins, G.S.; Riley, R.D.; Heinze, G.; Schuit, E.; Bonten, M.M.J.; Dahly, D.L.; Damen, J.A.; Debray, T.P.A.; et al. Prediction models for diagnosis and prognosis of covid-19 infection: Systematic review and critical appraisal. *BMJ* **2020**, *369*, m1328. [CrossRef]
58. Venkateswaran, N.; Galor, A.; Wang, J.; Karp, C.L. Optical coherence tomography for ocular surface and corneal diseases: A review. *Eye Vis.* **2018**, *5*, 13. [CrossRef]
59. Hall, M.A. *Correlation-Based Feature Selection for Machine Learning*; Department of Computer Science University of Waikato: Hamilton, New Zealand, 1999.
60. Bogunović, H.; Montuoro, A.; Baratsits, M.; Karantonis, M.G.; Waldstein, S.M.; Schlanitz, F.; Schmidt-Erfurth, U. Machine Learning of the Progression of Intermediate Age-Related Macular Degeneration Based on OCT Imaging. *Investig. Ophthalmol. Vis. Sci.* **2017**, *58*, BIO141–BIO150. [CrossRef]
61. Arcadu, F.; Benmansour, F.; Maunz, A.; Willis, J.; Haskova, Z.; Prunotto, M. Deep learning algorithm predicts diabetic retinopathy progression in individual patients. *NPJ Digit. Med.* **2019**, *2*, 92. [CrossRef]

62. Wang, M.; Shen, L.Q.; Pasquale, L.R.; Petrakos, P.; Formica, S.; Boland, M.; Wellik, S.R.; De Moraes, C.G.; Myers, J.S.; Saeedi, O.; et al. An Artificial Intelligence Approach to Detect Visual Field Progression in Glaucoma Based on Spatial Pattern Analysis. *Investig. Ophthalmol. Vis. Sci.* **2019**, *60*, 365–375. [[CrossRef](#)]
63. Bhuiyan, A.; Wong, T.Y.; Ting, D.S.W.; Govindaiah, A.; Souied, E.H.; Smith, R.T. Artificial Intelligence to Stratify Severity of Age-Related Macular Degeneration (AMD) and Predict Risk of Progression to Late AMD. *Transl. Vis. Sci. Technol.* **2020**, *9*, 25. [[CrossRef](#)]
64. Sample, P.A.; Boden, C.; Zhang, Z.; Pascual, J.; Lee, T.-W.; Zangwill, L.M.; Weinreb, R.N.; Crowston, J.G.; Hoffmann, E.M.; Medeiros, F.A.; et al. Unsupervised machine learning with independent component analysis to identify areas of progression in glaucomatous visual fields. *Investig. Ophthalmol. Vis. Sci.* **2005**, *46*, 3684–3692. [[CrossRef](#)]